

The number of graphs with large forbidden subgraphs

Béla Bollobás^{*†‡} and Vladimir Nikiforov^{†§}

January 24, 2010

Abstract

In this note, extending some results of Erdős, Frankl, Rödl, Alexeev, Bollobás and Thomason, we determine asymptotically the number of graphs which do not contain certain large subgraphs. In particular, we show that if H_1, H_2, \dots are graphs with $|H_n| = o(\log n)$ and $\chi(H_n) = r_n + 1$, then the number S_n of graphs of order n not containing H_n satisfies

$$\log_2 S_n = (1 - 1/r_n + o(1)) \binom{n}{2}.$$

We also give a similar statement for forbidden induced subgraphs.

Keywords: *number of graphs; induced subgraph; removal lemma;*

Introduction

Given a graph H , write $\mathcal{P}_n(H)$ for the set of all labelled graphs of order n not containing H . In 1976, Erdős, Kleitman and Rothschild [7] gave a theorem implying that

$$\log_2 |\mathcal{P}_n(K_{r+1})| = (1 - 1/r + o(1)) \binom{n}{2}. \quad (1)$$

In fact, this theorem, stated below, is considerably stronger than equation (1).

Theorem A *Given $r \geq 2$ and $\xi > 0$, there is $\rho = \rho(r, \xi)$ such that the number S_n of labelled graphs of sufficiently large order n containing at most ρn^{r+1} copies of K_{r+1} satisfies*

$$(1 - 1/r) \binom{n}{2} \leq \log_2 S_n \leq (1 - 1/r + \xi) \binom{n}{2}.$$

^{*}Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge CB3 0WB, UK and

[†]Department of Mathematical Sciences, University of Memphis, Memphis TN 38152, USA

[‡]Research supported in part by NSF grants DMS-0505550, CNS-0721983, CCF-0728928 and DMS-0906634, and ARO grant W911NF-06-1-0076

[§]Research supported by NSF Grant DMS-0906634.

Ten years later, Erdős, Frankl and Rödl [8] showed that the conclusion in (1) holds if K_{r+1} is replaced by an arbitrary fixed $(r+1)$ -chromatic graph H . Surprisingly, Theorem A, combined with an observation Erdős [6] made in 1964 (Theorem E below), easily imply the result of Erdős, Frankl and Rödl, and even stronger ones: for details see the first concluding remark at the end.

Here we shall give essentially best possible results that can be obtained by replacing H with a sequence of forbidden graphs whose order grows with n . More precisely, given integers $r \geq 2$, $p \geq 1$, $q \geq 1$ and real $c \in (0, 1/2)$, write $K_{r+1}(p; q)$ for the complete $(r+1)$ -partite graph with r parts of size p and one part of size q . Here and further, \log with unspecified base stands for the natural logarithm.

Our first result is the following theorem.

Theorem 1 *Given $r \geq 2$ and $0 < \varepsilon \leq 1/2$, there is $\delta = \delta(\varepsilon) > 0$ such that for n sufficiently large,*

$$(1 - 1/r) \binom{n}{2} \leq \log_2 \left| \mathcal{P}_n \left(K_{r+1} \left(\lfloor \delta \log n \rfloor ; \lceil n^{1-\sqrt{\delta}} \rceil \right) \right) \right| \leq (1 - 1/r + \varepsilon) \binom{n}{2}.$$

As we shall see, when ε decreases, so does $\delta = \delta(\varepsilon)$; this has the somewhat peculiar consequence that when ε decreases, the order of the forbidden graph $K_{r+1} \left(\lfloor \delta \log n \rfloor ; \lceil n^{1-\sqrt{\delta}} \rceil \right)$ increases; in fact, with the function $\delta(\varepsilon)$ we shall take, this order is $\Theta(n^{1-o(\varepsilon)})$.

Forgetting most of the vertices in the large vertex class of $K_{r+1} \left(\lfloor \delta \log n \rfloor ; \lceil n^{1-\sqrt{\delta}} \rceil \right)$, we get the following simplified assertion.

Corollary 2 *Let (H_n) be a sequence of graphs, with $|H_n| = o(\log n)$ and $\chi(H_n) = r_n + 1$. Then, for every $\varepsilon > 0$ and n large enough,*

$$(1 - 1/r_n) \binom{n}{2} \leq \log_2 |\mathcal{P}_n(H_n)| \leq (1 - 1/r_n + \varepsilon) \binom{n}{2}.$$

Indeed, if $r_n > 1/\varepsilon$, there is nothing to prove. If $r_n < 1/\varepsilon$ then, as H_n is a subgraph of the complete $(r_n + 1)$ -partite graph with all parts of size $|H_n| = o(\log n)$, the upper bound follows when n is sufficiently large. The lower bound follows as in Theorem 1.

We should like to emphasize that Szemerédi's Regularity Lemma, a standard tool to tackle questions like this, will not be used in our proof of Theorem 1, not even indirectly.

Next we turn to forbidden induced subgraphs, where the role of the chromatic number is played by the *coloring number* χ_c , defined first in [4], and given below.

Definition 3 *Let $0 \leq s \leq r$ be integers and let $\mathcal{H}(r, s)$ be the class of graphs whose vertex sets can be partitioned into s cliques and $r - s$ independent sets. Given a graph property \mathcal{P} , the coloring number $\chi_c(\mathcal{P})$ is defined as*

$$\chi_c(H) = \max \{r : \mathcal{H}(r, s) \subset \mathcal{P} \text{ for some } s \in [r]\}$$

Also, given a graph H , let us write $\mathcal{P}_n^*(H)$ for the set of graphs of order n not containing H as an induced subgraph; clearly $\mathcal{P}_n^*(H)$ is a hereditary property.

A special case of a general result proved by Alexeev [1] and independently by Bollobás and Thomason [4],[5] is the exact analogue of (1): if H is a fixed graph and $r = \chi_c(\mathcal{P}_n^*(H))$, then

$$\log_2 |\mathcal{P}_n^*(H)| = (1 - 1/r + o(1)) \binom{n}{2}. \quad (2)$$

Motivated by Theorem A, we first observe the following assertion, which is an immediate consequence of the removal lemma of Alon, Fisher, Krivelevich and Szegedy [3] (Theorem B below) and the Alexeev-Bollobás-Thomason result (2). We state it as a theorem only to properly match Theorem A for induced graphs.

Theorem 4 *Let H be a graph and let $r = \chi_c(\mathcal{P}_n^*(H))$. For every $\xi > 0$, there is a $\rho = \rho(H, \xi)$ such that the number S_n of graphs of sufficiently large order n containing at most $\rho n^{|H|}$ induced copies of H satisfies*

$$(1 - 1/r) \binom{n}{2} \leq \log_2 S_n \leq (1 - 1/r + \xi) \binom{n}{2}.$$

Note that our proof of this theorem uses implicitly Szemerédi's Regularity Lemma. It would be interesting to find a proof avoiding this lemma. We know from Erdős, Kleitman and Rothschild [7] that this can be done when $H = K_{r+1}$.

Next, we shall show that the conclusion in (2) holds when H is replaced by a sequence of forbidden graphs whose order grows with n . To give the precise statement, we need the following definition.

Definition 5 *Given a labelled graph H with $V(H) = [h]$ and positive integers p_1, \dots, p_h , we say that a graph F is of type $H(p_1, \dots, p_h)$ if F can be obtained by replacing each vertex $u \in V(H)$ with a graph G_u of order p_u and each edge $uv \in E(H)$ with a complete bipartite graph with vertex classes $V(G_u)$ and $V(G_v)$; if $uv \notin E(H)$ and $u \neq v$, then F contains no edges between $V(G_u)$ and $V(G_v)$.*

Now, given a labelled graph H and positive integers p and q , let

$$\mathcal{P}_n(H; p, q) = \left\{ G : \begin{array}{l} G \text{ is a labelled graph of order } n \text{ and } G \text{ contains no induced} \\ \text{subgraph of type } H(p, \dots, p, q) \end{array} \right\}.$$

Here is our second main result.

Theorem 6 *Let H be a labelled graph and let $r = \chi_c(\mathcal{P}_n^*(H))$. For every $\varepsilon > 0$, there is $\delta = \delta(\varepsilon) > 0$ such that for n sufficiently large*

$$(1 - 1/r) \binom{n}{2} \leq \log_2 \left| \mathcal{P}_n \left(H; \lfloor \delta \log n \rfloor, \lceil n^{1-\sqrt{\delta}} \rceil \right) \right| \leq (1 - 1/r + \varepsilon) \binom{n}{2}. \quad (3)$$

In some sense Theorems 1 and 6 are almost best possible, in view of the following simple observation, that can be proved by considering the random graph $G_{n,p}$ with $p \rightarrow 1$.

Given $r \geq 2$ and $\varepsilon > 0$, there is $C > 0$ such that the number of graphs S_n which do not contain $K_2(\lceil C \log n \rceil, \lceil C \log n \rceil)$ satisfies $S_n \geq (1 - \varepsilon) 2^{\binom{n}{2}}$.

Proofs

For the proof of Theorem 4 we need a version of the Removal Lemma of Ruzsa and Szemerédi [11] for induced graphs; this result was stated and proved in [3].

Theorem B *Given a graph H and $\alpha > 0$, there is $\beta = \beta(\alpha) > 0$ such that if a graph G of order n contains fewer than $\beta n^{|H|}$ induced copies of H , then one can change at most αn^2 edges of G so that the resulting graph does not contain an induced copy of H .*

We need also the following facts, which are Theorem 1 of [9] and Theorem 2 of [10].

Theorem C *Let $r \geq 3$, $(\ln n)^{-1/r} \leq c \leq 1/2$, and let G be a graph with n vertices. If G contains more than cn^r copies of K_r , then G contains a $K_r(s, \dots, s, t)$ with $s = \lfloor c^r \ln n \rfloor$ and $t > n^{1-c^{r-1}}$.*

Theorem D *Let $2 \leq h \leq n$, $(\ln n)^{-1/h^2} \leq c \leq 1/4$, let H be a graph of order h , and G be a graph of order n . If G contains more than cn^h induced copies of H , then G contains an induced subgraph of type $H(s, \dots, s, t)$, where $s = \lfloor c^{h^2} \ln n \rfloor$ and $t > n^{1-c^{h-1}}$.*

Note that Theorems 1, 6, and 4 have to be proved for n sufficiently large; thus, in the proofs below, we shall assume that n is as large as needed.

Proof of Theorem 1. Write $T_r(n)$ for the r -partite Turán graph of order n and note that no subgraph of $T_r(n)$ contains a K_{r+1} . Also, note that the number s'_n of labelled spanning subgraphs of $T_r(n)$ satisfies

$$\log_2 s'_n \geq (1 - 1/r) \frac{n^2}{2} - \frac{r}{8} \geq (1 - 1/r) \binom{n}{2},$$

proving the lower bound in (3); thus, to finish the proof of Theorem 1 we need to prove the upper bound in (3).

Fix $\varepsilon > 0$, let $\rho(r, \cdot)$ be the function from Theorem A, and set $\delta = \rho(r, \varepsilon)^{r+1}$. If a graph G does not contain a $K_{r+1}(\lfloor \delta \log n \rfloor; \lceil n^{1-\sqrt{\delta}} \rceil)$, then Theorem C implies that G contains at most $\delta^{1/(r+1)} n^{r+1} = \rho(r, \varepsilon) n^{r+1}$ copies of K_{r+1} ; in turn, Theorem A implies that

$$\log_2 \left| \mathcal{P}_n \left(K_{r+1} \left(\lfloor \delta \log n \rfloor; \lceil n^{1-\sqrt{\delta}} \rceil \right) \right) \right| \leq (1 - 1/r + \varepsilon) \binom{n}{2},$$

completing the proof of Theorem 1. □

Proof of Theorem 4. The lower bound is immediate since S_n must be at least as large as the number of graphs in $\mathcal{H}(r, s)$ of order n , and so, as in Theorem 1, we see that

$$\log_2 S_n \geq (1 - 1/r) \frac{n^2}{2} - \frac{r}{8} \geq (1 - 1/r) \binom{n}{2}.$$

Let us now prove the upper bound. Fix $\varepsilon > 0$, and let σ be such that

$$\frac{\varepsilon}{3} \geq \sigma \log_2 \frac{4}{\sigma}.$$

Let $\beta(\cdot)$ be the function of Theorem B, and set $\delta = \beta(\sigma/2)$. If a graph G of order n contains at most δn^h induced copies of H , then Theorem B implies that all induced copies of H in G can be destroyed by changing at most $(\sigma/2)n^2$ edges. Therefore, we see that

$$\begin{aligned} \log_2 S_n &\leq \log_2 |\mathcal{P}_n^*(H)| + \log_2 \binom{\binom{n}{2}}{\sigma n^2/2} \leq \log_2 p_n + \sigma \frac{n^2}{2} \log_2 \frac{4}{\sigma} \leq \log_2 p_n + \frac{\varepsilon}{3} \frac{n^2}{2} \\ &\leq \log_2 p_n + \frac{\varepsilon}{2} \binom{n}{2} \leq (1 - 1/r + \varepsilon) \binom{n}{2}. \end{aligned}$$

The last inequality above follows from the Alexeev-Bollobás-Thomason result. This completes the proof of Theorem 4. \square

Proof of Theorem 6. The lower bound is determined as in Theorem 4, so let us prove the upper bound. Let $\rho(H, \cdot)$ be the function from Theorem 4. Fix $\varepsilon > 0$, let $\delta = \rho(H, \varepsilon)^{h^2}$, and set $p = \lfloor \delta \log n \rfloor$, $q = \lceil n^{1-\sqrt{\delta}} \rceil$. Suppose that a graph G of order n contains no induced subgraph of type $H(p, \dots, p, q)$. Then, Theorem D implies that G contains at most $\delta^{1/h^2} n^h = \rho(H, \varepsilon) n^h$ induced copies of H . In turn, Theorem 4 implies that the number S_n of such graphs satisfies

$$\log_2 S_n \leq (1 - 1/r + \varepsilon) \binom{n}{2},$$

completing the proof of Theorem 6. \square

Concluding remarks

1. As mentioned at the beginning of this note, in [6], equation (18'), Erdős gave a result about uniform hypergraphs, which implies the following statement about graphs:

Theorem E *Let $r \geq 2$. If a graph G of order n contains at least εn^r copies of K_r , then G contains a copy of*

$$K_r \left(\left\lfloor \delta (\log n)^{1/(r-1)} \right\rfloor, \dots, \left\lfloor \delta (\log n)^{1/(r-1)} \right\rfloor \right)$$

for some $\delta = \delta(\varepsilon) > 0$.

In view of Theorem A, we immediately see the following corollary:

Given $r \geq 2$ and $\varepsilon > 0$, there is $\delta = \delta(\varepsilon)$ such that for n sufficiently large,

$$\log_2 \left| \mathcal{P}_n \left(K_{r+1} \left(\left\lfloor \delta (\log n)^{1/r} \right\rfloor, \dots, \left\lfloor \delta (\log n)^{1/r} \right\rfloor \right) \right) \right| \leq (1 - 1/r + \varepsilon) \binom{n}{2}. \quad (4)$$

This statement could have been published as early as 1976, but the authors of Theorem A somehow missed it, albeit Theorem E was used indeed in the proof of Theorem A (see [7], p. 20, line -5).

2. It is possible that the approach of [9] can give an explicit expression for $\delta(\varepsilon)$ in Theorem 1. This would help one to estimate how much $\left| \mathcal{P}_n \left(K_{r+1} \left(\lfloor \delta \log n \rfloor; \lceil n^{1-\sqrt{\delta}} \rceil \right) \right) \right|$ is larger than the number of r -partite graphs of order n .

3. We reiterate the problem mentioned above: prove Theorem 4 avoiding the use of Szemerédi’s Regularity Lemma.

4. In the last two decades, the study of the number of graphs with given properties has acquired a truly remarkable scale and sophistication, see, e.g., [2] and its references. Yet, we do not see a simple way to accommodate the above results within this general framework.

References

- [1] V.E. Alekseev, On the entropy values of hereditary classes of graphs, *Discrete Math. Appl.* **3** (1993), 191–199.
- [2] N. Alon, J. Balogh, B. Bollobás and R. Morris, The structure of all graphs in a hereditary property, preprint available at *2009arXiv0905.1942A*
- [3] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs, *Combinatorica*, **20** (2000) 451–476.
- [4] B. Bollobás and A. Thomason, Projections of bodies and hereditary properties of hypergraphs, *Bull. London Math. Soc.* **27** (1995), 417–424.
- [5] B. Bollobás and A. Thomason, Hereditary and monotone properties of graphs, “The mathematics of Paul Erdős, II” (R.L. Graham and J. Nešetřil, Editors), *Alg. and Combin.*, Vol. 14, Springer-Verlag, New York/Berlin (1997), 70–78.
- [6] P. Erdős, On extremal problems of graphs and generalized graphs, *Israel J. Math.* **2** (1964), 183–190.
- [7] P. Erdős, D.J. Kleitman and B.L. Rothschild, Asymptotic enumeration of K_n -free graphs, in *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973)*, Vol. II, pp. 19–27. Atti dei Convegni Lincei, 17, Accad. Naz. Lincei, Roma, 1976.
- [8] P. Erdős, P. Frankl and V. Rödl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, *Graphs Combin.* **2** (1986), 113–121.
- [9] V. Nikiforov, Graphs with many r -cliques have large complete r -partite subgraphs, *Bull. London Math. Soc.* **40** (2008), 23–25.
- [10] V. Nikiforov, Graphs with many copies of a given subgraph, *Electronic J. Combin.* **15** (2008), N6.
- [11] I. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, in *Combinatorics (Keszthely, 1976)*, *Coll. Math. Soc. J. Bolyai* 18, Vol. II, 939–945.